Module 6: Introduction to Machine Learning with Python

Case Study

Case Study

Objectives:

1.Provide the learner some more practice for exploratory data analysis.

2.Equip the learner to fit and evaluate a linear regression model.

Questions:

1.Load the data from "cereal.csv" and plot histograms of sugar and vitamin content across different cereals.

[Hint: Extract values of a specific column using their labels and use hist method of pyplot]

2. The names of the manufactures are coded using alphabets, create a new column with their full name using the below mapping.

'N': 'Nabisco',

'Q': 'Quaker Oats',

'K': 'Kelloggs',

'R': 'Raslston Purina',

'G': 'General Mills' ,

'P' :'Post' ,

'A':'American Home Foods Products'

Create a bar plot where each manufacturer is on the y axis and the height of the bars depict the number of cereals manufactured by them.

[Hint: Try using countplot this time or bar method of pyplot]

3. Extract the rating as your target variable ‘y’ and all numerical parameters as your predictors ‘x’. Separate 25% of your data as test set.

4. Fit a linear regression module and measure the mean squared error on test dataset.

[Hint: Explore linear models and metrics section of sklearn documentation]